

Identification of DSGE Models - the Effect of Higher-Order Approximation and Pruning ^{*}

Willi Mutschler[†]

*Institute for Econometrics and Economic Statistics,
Center for Quantitative Economics (CQE),
University of Münster, Germany*

Version: August 1, 2014

CQE Working paper.

Comments are greatly appreciated.

Abstract

Several formal methods have been proposed to check local identification in DSGE models using rank criteria. However, all methods focus on the linear solution of the DSGE model to the first order. Recently there has been huge progress in the estimation of non-linear DSGE models, yet formal identification criteria are missing. The contribution of the paper is threefold: First, we extend the existent methods to higher-order approximations and establish rank criteria for local identification given the pruned state-space representation. It is shown that this may improve overall identification of a DSGE model via imposing additional restrictions on the moments and spectrum. Second, we derive analytical derivatives for the reduced-form matrices, unconditional moments and spectral density for the pruned state-space system. Third, using a second-order approximation, we are able to identify previously non-identifiable parameters: namely the parameters govern the investment adjustment costs in the Kim (2003) model and the coefficients of the Taylor-rule in the An and Schorfheide (2007) model.

JEL Classification: C10, C51, C52, E1

^{*}This paper is an improved version of some sections in an earlier paper entitled: "Identification of DSGE Models - A Comparison of Methods and the Effect of Second-Order Approximation", which was presented at the course on Identification Analysis and Global Sensitivity Analysis for Macroeconomic Models in Ispra, Italy (2013); at the Fall Macro Midwest Meeting in Minneapolis, USA (2013); at the Spring Meeting of Young Economists in Vienna, Austria (2014); at the Workshop on Empirical Methods in Macroeconomic Policy Analysis in Bucharest, Romania (2014); at the Asian Meeting of the Econometric Society in Taipei, Taiwan (2014); at the European Meeting of the Econometric Society in Toulouse, France (2014); at the Annual Conference of the Verein für Socialpolitik in Hamburg, Germany (2014); and at the DYNARE Conference in Paris, France (2014). The author thanks Nikolay Iskrev, Michel Juillard, Junior Maih, Zhongjun Qu, Marco Ratto, Stephanie Schmitt-Grohé, Ron Smith, Mark Trede, Bernd Wilfing and seminar participants for critical and helpful comments and encouragement.

[†]Corresponding author. Center for Quantitative Economics, Institut für Ökonometrie und Wirtschaftsstatistik, Westfälische Wilhelms-Universität Münster, Am Stadtgraben 9, 48143 Münster, Tel.: +49-251-83-22914, Fax: +49-251-83-22012, Email: willi.mutschler@uni-muenster.de.

1 Introduction

Many different methods for solving and estimating DSGE models have been developed and used in order to get a detailed analysis and thorough estimation of dynamic macroeconomic relationships. Recently, the question of identification of DSGE models has proven to be of major importance, especially since identification of a model precedes estimation and inference. Several formal methods have been proposed to check local identification in linearized DSGE models via rank criteria (Iskrev 2010; Komunjer and Ng 2011; Qu and Tkachenko 2012) or Bayesian indicators (Koop, Pesaran, and Smith 2013). Whereas there is a growing literature on non-linear estimation of DSGE models (Andreasen 2011; Andreasen 2013; Fernández-Villaverde and Rubio-Ramírez 2007; Herbst and Schorfheide 2013; Ivashchenko 2014; Kollman 2014), all identification methods focus on the linear approximation of the DSGE model to the first order. In this paper we will relax this assumption and establish rank criteria for non-linear DSGE models solved by higher-order approximation of the policy functions. Intuitively, this may yield additional restrictions on the moments and spectrum of the model that can be used to identify previously unidentified (sets of) parameters.

However, there is a caveat, since higher-order approximations sometimes yield explosive or non-stationary processes. Therefore, we use the pruning scheme proposed by Kim et al. (2008), who show that the pruned state-space is stationary and ergodic. Further Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2014) derive closed-form expressions for unconditional moments up to third-order approximations. Exploiting these results, the contribution of this paper is threefold. First, we show how to extend the existent identification criteria based on ranks for higher-order approximations. Throughout the exposition we focus on a second-order approximation, since extending ideas and propositions is – apart from notation – conceptually straightforward for higher-order approximations. Second, we show how to analytically calculate the Jacobians of the mean, autocovariogram, and the spectrum of the pruned state-space w.r.t. the deep parameters of the model. Third, to make our exposition illustrative, all methods are applied on two models that are known to have lack of identification in their (log-)linearized versions: the Kim (2003) and the An and Schorfheide (2007) model. In particular, we show that the parameters governing the adjustment costs in Kim (2003) as well as the coefficients of the Taylor-rule in An and Schorfheide (2007) can be identified using a second-order approximation and the pruned state-space.

The ideas and procedures derived are useful both from a theoretical and applied point of view. Theoretically, this paper adds to the literature on local identification of non-linear DSGE models by establishing rank criteria and analytic derivatives for higher-order approximations using the pruned state-space representation. Based on these findings, we believe that the suggested approach is a useful new tool before actually taking non-linear DSGE models to data. In particular, an applied researcher can check whether unidentified parameters may be estimable using higher-order approximations even before she actually uses tedious non-linear estimation methods. Our Matlab-code is model-independent and can be found on the homepage of the author.

2 DSGE framework

Let E_t be the expectation operator conditional on information available at time t , then

$$\begin{aligned} 0 &= E_t f(x_{t+1}, y_{t+1}, x_t, y_t | \theta), \\ x_{t+1} &= h(x_t, \sigma | \theta) + \sigma \eta(\theta) \varepsilon_{t+1}, \\ y_t &= g(x_t, \sigma | \theta) \end{aligned}$$

is called the general DSGE model with deep parameters θ , states x_t , controls y_t , observables d_t , stochastic innovations ε_t , and perturbation parameter σ , which can be cast into a non-linear first-order system of expectational difference equations f . For the sake of notation, we assume that all control variables are observable. Further, ε_t is iid with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_t') = I_{n_\varepsilon}$; thus, the variances and cross-correlations are captured by the premultiplied matrix η . σ is set to be dependent on the standard deviation of one of the shocks, while scaling all other variances and cross-correlations in η accordingly. See appendix E on how to squeeze the example models into this framework.¹

The solution of such rational expectation models is characterized by so-called *policy-functions*, g and h , that solve (at least approximately) the system of equations f . We follow Schmitt-Grohé and Uribe (2004) and use perturbation techniques to solve the model around the non-stochastic steady-state given by $\bar{x} = h(\bar{x}, 0 | \theta)$, $\bar{y} = g(\bar{x}, 0 | \theta)$ and $f(\bar{x}, \bar{y}, \bar{x}, \bar{y} | \theta) = 0$. Moreover, we exploit ideas of Gomme and Klein (2011) to approximate the policy functions using the Magnus and Neudecker (1999) definition of the Hessian.² Denote the Jacobian of f evaluated at the steady-state as $\mathcal{D}f(\bar{x}, \bar{y}) := \left(\frac{\partial f(\bar{x}, \bar{y})}{\partial x'_{t+1}} \quad \frac{\partial f(\bar{x}, \bar{y})}{\partial y'_{t+1}} \quad \frac{\partial f(\bar{x}, \bar{y})}{\partial x'_t} \quad \frac{\partial f(\bar{x}, \bar{y})}{\partial y'_t} \right) := \begin{pmatrix} f_1 & f_2 & f_3 & f_4 \end{pmatrix}$, then

$$\mathcal{H}f(\bar{x}, \bar{y}) := \mathcal{D}vec([\mathcal{D}f(\bar{x}, \bar{y})]')$$

is defined as the Magnus-Neudecker Hessian of f evaluated at the non-stochastic steady-state. This definition simplifies the computations as well as the analytical derivatives, since no tensor notation is needed and basic matrix algebra can be used, see appendix A for further reference.³ The second-order Taylor approximation at the non-stochastic steady-state is then given by

$$\hat{x}_{t+1} = h_x \cdot \hat{x}_t + \frac{1}{2} [I_{n_x} \otimes \hat{x}'_t] \cdot h_{xx} \cdot \hat{x}'_t + \frac{1}{2} \sigma^2 h_{\sigma\sigma} + \sigma \eta \varepsilon_{t+1}, \quad (1)$$

$$\hat{y}_t = g_x \cdot \hat{x}_t + \frac{1}{2} [I_{n_y} \otimes \hat{x}'_t] \cdot g_{xx} \cdot \hat{x}'_t + \frac{1}{2} \sigma^2 g_{\sigma\sigma}, \quad (2)$$

where $\hat{x}_{t+1} = x_{t+1} - \bar{x}$ denotes deviations from steady-state. Further g_x and g_σ are the gradients of g with respect to x_t and σ respectively, g_{xx} and $g_{\sigma\sigma}$ the corresponding Magnus-Neudecker-Hessians,

¹The state and control vectors can be handled very flexible depending on the model. In particular, in the proposed setup innovations enter only linearly into the framework, whereas e.g. in the Dynare framework they may also enter nonlinearly. However, by simply appending auxiliary equations to the model and extending the state vector both representations are equivalent, see the technical appendix of Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2014). In the same fashion, we are able to add measurement equations and measurement errors by simply extending the states and controls. For instance, in our Matlab code we append auxiliary variables for our observables to the control vector and then premultiply a known selection matrix to get the policy functions of observable variables.

²For a third-order approximation using Magnus-Neudecker Hessians see Binning (2013).

³For recent literature in favor of this definition see also Magnus (2010) and Pollock (2013).

all evaluated at $(\bar{x}, 0)$. The same notation applies to $\hat{y}_t, h_x, h_\sigma, h_{xx}$ and $h_{\sigma\sigma}$. Schmitt-Grohé and Uribe (2004) show that all linear terms as well as cross-terms in σ , i.e. $g_\sigma, g_{x\sigma}, g_{\sigma x}, h_\sigma, h_{x\sigma}, h_{\sigma x}$, are equal to zero, since the approximation is around $\sigma = 0$. Notice also, that in a linearization to the first-order (or log-linearization) all terms lead by 1/2 drop out.

There are several methods and algorithms to calculate the matrices h_x and g_x , since these are the coefficients of a first-order linearization or log-linearization of the model. We follow Klein (2000) to obtain h_x and g_x using the generalized Schur decomposition.⁴ The other matrices can be calculated by inserting the policy functions into the model equations

$$E_t f(h(x_t, \sigma) + \sigma \eta \varepsilon_{t+1}, g(h(x_t, \sigma) + \sigma \eta \varepsilon_{t+1}, \sigma), x_t, g(x_t, \sigma)) = 0.$$

This expression is known at the non-stochastic steady-state; thus, all derivatives of f must be 0 when evaluated at the steady-state. Differentiating f twice using the chain-rule of Magnus and Neudecker (1999, p. 110), evaluating the Jacobian $\mathcal{D}f = (f_1 \ f_2 \ f_3 \ f_4)$ and Hessian H of f at the non-stochastic steady-state, and setting it to 0 yields (after some algebra):

$$\begin{bmatrix} \text{vec}(g_{xx}) \\ \text{vec}(h_{xx}) \end{bmatrix} = -Q^{-1} \text{vec}(R), \quad \begin{bmatrix} h_{ss} \\ g_{ss} \end{bmatrix} = -S^{-1}U, \quad (3)$$

with

$$\begin{aligned} Q &= \begin{bmatrix} h'_x \otimes f_2 \otimes h'_x + I_{n_x} \otimes f_4 \otimes I_{n_x} & I_{n_x} \otimes (f_1 \otimes I_{n_x} + f_2 g_x \otimes I_{n_x}) \\ & \end{bmatrix}, \\ R &= (I_{n_x+n_y} \otimes M')HM, \\ S &= \begin{bmatrix} f_1 + f_2 g_x & f_2 + f_4 \end{bmatrix}, \\ U &= f_2 \text{trm}[(I_{n_y} \otimes (\eta \eta'))g_{xx}] + \text{trm}[(I_{n_x+n_y} \otimes N')HN(\eta \eta')], \\ M &= \begin{bmatrix} h_x \\ g_x h_x \\ I_{n_x} \\ g_x \end{bmatrix}, \quad N = \begin{bmatrix} I_{n_x} \\ g_x \\ 0_{(n_x+n_y) \times n_x} \end{bmatrix}, \end{aligned}$$

and trm defines the matrix trace of an $nm \times n$ matrix $[Y'_1 \ Y'_2 \ \dots \ Y'_m]'$ as the $m \times 1$ vector $[\text{tr}(Y_1) \ \text{tr}(Y_2) \ \dots \ \text{tr}(Y_m)]'$. See Gomme and Klein (2011) for the derivation. For our purpose it is sufficient to note that there exist analytical closed-form solutions that we will differentiate with respect to the deep parameters in section 4.

3 Pruned state-space system

The approximations (1) and (2) are a straightforward application of Taylor series expansions to the state variables. However, simulation studies show that due to artificial fixed points, higher-order approximations often generate explosive time-paths even though the linear approximation is stable.

⁴See Anderson (2008) for a comparison of algorithms, which are basically all equivalent and differ only (slightly) in computational burden. Further, all provide and check the Blanchard and Kahn (1980) conditions that are necessary in order to have a stable saddle-path solution, i.e. a unique mapping between state and control variables.

Thus, the model may neither be stationary nor imply an ergodic probability distribution, both assumptions are essential for identification and estimation purposes. Thus, Kim et al. (2008) propose the pruning scheme, in which one leaves out terms in the solution that have higher-order effects than the approximation order. For instance, given a second-order approximation, we decompose the state vector into first-order (\hat{x}^f) and second-order (\hat{x}^s) effects, and set up the law of motions for these variables preserving only effects up to second-order (see Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2014) for details):

$$\hat{x}_{t+1}^f = h_x \hat{x}_t^f + \sigma \eta \varepsilon_{t+1} \quad (4)$$

$$\hat{x}_{t+1}^s = h_x \hat{x}_t^s + \frac{1}{2} H_{xx} (x_t^f \otimes x_t^f) + \frac{1}{2} h_{\sigma\sigma} \sigma^2 \quad (5)$$

$$\hat{x}_{t+1} = \hat{x}_{t+1}^f + \hat{x}_{t+1}^s = h_x (\hat{x}_t^f + \hat{x}_t^s) + \frac{1}{2} H_{xx} (\hat{x}_t^f \otimes \hat{x}_t^f) + \frac{1}{2} h_{\sigma\sigma} \sigma^2 + \sigma \eta \varepsilon_{t+1} \quad (6)$$

$$\hat{y}_t = \hat{y}_t^f + \hat{y}_t^s = g_x (\hat{x}_t^f + \hat{x}_t^s) + \frac{1}{2} G_{xx} (\hat{x}_t^f \otimes \hat{x}_t^f) + \frac{1}{2} g_{\sigma\sigma} \sigma^2 \quad (7)$$

with H_{xx} being an $n_x \times n_x^2$ matrix containing all second-order terms for the i -th state variable in the i -th row, whereas G_{xx} is an $n_y \times n_x^2$ matrix containing all second-order terms for the i -th control variable in the i -th row.⁵ Thus, terms containing $\hat{x}_t^f \otimes \hat{x}_t^s$ and $\hat{x}_t^s \otimes \hat{x}_t^s$ are left out, since they reflect third-order and fourth-order effects which are higher than the approximation order.⁶

It is convenient to extend the state vector to $z_t := [(\hat{x}_t^f)', (\hat{x}_t^s)', (\hat{x}_t^f \otimes \hat{x}_t^f)']'$, then equations (4), (5), (6) and (7) can be rewritten as a linear system of equations called the pruned state-space representation:

$$z_{t+1} = Az_t + B\xi_{t+1} + c \quad (8)$$

$$y_t = \bar{y} + Cz_t + d \quad (9)$$

where

$$A := \begin{bmatrix} h_x & 0 & 0 \\ 0 & h_x & \frac{1}{2} H_{xx} \\ 0 & 0 & h_x \otimes h_x \end{bmatrix}, \quad B := \begin{bmatrix} \sigma \eta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \sigma \eta \otimes \sigma \eta & \sigma \eta \otimes h_x & h_x \otimes \sigma \eta \end{bmatrix}, \quad C := \begin{bmatrix} g_x & g_x & \frac{1}{2} G_{xx} \end{bmatrix},$$

$$\xi_{t+1} := \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_{t+1} \otimes \varepsilon_{t+1} - \text{vec}(I_{n_\varepsilon}) \\ \varepsilon_{t+1} \otimes x_t^f \\ x_t^f \otimes \varepsilon_{t+1} \end{bmatrix}, \quad c := \begin{bmatrix} 0 \\ \frac{1}{2} h_{\sigma\sigma} \sigma^2 \\ (\sigma \eta \otimes \sigma \eta) \text{vec}(I_{n_\varepsilon}) \end{bmatrix}, \quad d := \frac{1}{2} g_{\sigma\sigma} \sigma^2.$$

Thus, conceptually we work in a state-space system with a linear law of motion in z_t that is very similar to a log-linearized DSGE model; hence, many concepts simply carry over.⁷ For instance, it

⁵ H_{xx} and G_{xx} can be obtained from h_{xx} and g_{xx} respectively using Matlab's permute and reshape functions.

⁶This may seem an ad-hoc procedure, however, pruning can also be theoretically founded as a Taylor expansion in the perturbation parameter (Lombardo and Uhlig 2014) or on an infinite moving average representation (Lan and Meyer-Gohde 2013).

⁷This approach also works for higher-order approximations: Appending the state vector accordingly, we are able to establish a similar linear system in an extended state vector.

can be shown that if the first-order approximation is stable, i.e. all Eigenvalues of h_x have modulus less than one, then the pruned state-space is also stable, i.e. all higher-order terms are unique and all Eigenvalues of A have modulus less than one. Further if ε_t has finite fourth moments, then the pruned state-space system has finite second moments.⁸ Standard results from VAR(1) systems can be thus used regarding the computation of unconditional moments and spectrum.

First, it is trivial to show that ξ_t is iid with $E(\xi_t) = 0$ and finite covariance matrix $\Sigma_\xi := E(\xi_t \xi_t')$, since it is a function of x_t^f, ε_{t+1} and $\varepsilon_{t+1} \otimes \varepsilon_{t+1}$.⁹ The mean of the extended state vector is equal to

$$\mu_z := E(z_t) = (I_{2n_x+n_z^2} - A)^{-1}c. \quad (10)$$

Intuitively the mean of the pruned state-space consists of two effects: The first-order effect ($E(\hat{x}_t^f) = 0$) simply states certainty-equivalence, i.e. the mean of x_t is equal to the steady-state in a first-order approximation. Using a second-order approximation we adjust the mean for risk given a constant $\frac{1}{2}h_{\sigma\sigma}\sigma^2$ and the variance of the states $vec(\Sigma_x) := E(\hat{x}_t^f \otimes \hat{x}_t^f) = (I_{n_x^2} - h_x)^{-1}(\sigma\eta_x \otimes \sigma\eta_x)vec(I_{n_x})$.

For the covariance matrix $\Sigma_z := E[(z_t - \mu_z)(z_t - \mu_z)']$, we have

$$\Sigma_z = A\Sigma_z A' + B\Sigma_\xi B'. \quad (11)$$

Using an algorithm for Lyapunov equations or vectorization we can solve (11)

$$vec(\Sigma_z) = (I_{(2n_x+n_z^2)^2} - A \otimes A)^{-1}vec(B\Sigma_\xi B')$$

and are hence able to calculate the autocovariances for $t \in \mathbb{N} \setminus \{0\}$:

$$\Sigma_z(t) := E[(z_t - E(z_t))(z_0 - E(z_0))'] = A^t \Sigma_z. \quad (12)$$

Since there is a linear relationship between y_t and z_t in (9), we get closed-form expressions for the unconditional moments of our controls. That is, for $t \in \mathbb{N} \setminus \{0\}$

$$\mu_y := E(y_t) = \bar{y} + C\mu_z + d, \quad (13)$$

$$\Sigma_y := E[(y_t - \mu_y)(y_t - \mu_y)'] = C\Sigma_z C', \quad (14)$$

$$\Sigma_y(t) := E[(y_t - \mu_y)(y_0 - \mu_y)'] = C\Sigma_z(t)C'. \quad (15)$$

For the spectral density consider the vector-moving-average representation (VMA) of states

$$z_t = \mu_z + \sum_{j=0}^{\infty} A^j B \xi_{t-j}.$$

⁸This is basically Proposition 1 in Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2014).

⁹ Σ_ξ can be partitioned into several submatrices which can be computed element-by-element. Our code is based on the computationally more efficient method to compute $B\Sigma_\xi B'$, see the technical appendix of Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2014).

Using equation (9) and lag-operator L , we thus get the VMA for our controls

$$y_t - \bar{y} - C\mu_z - d = \sum_{j=0}^{\infty} CA^j B\xi_{t-j} = H_\xi(L)B\xi_t$$

with transfer function $H_\xi(L) = C(I_{(2n_x+n_z)^2} - AL)^{-1}$. Using the Fourier transformation for the lag-operator L the spectral density matrix Ω_y is given by

$$\Omega_y(\omega) = \frac{1}{2\pi} H_\xi(e^{-i\omega}) \cdot B\Sigma_\xi B' \cdot H_\xi(e^{-i\omega})^*, \quad \omega \in [-\pi; \pi], \quad (16)$$

with $*$ denoting the conjugate transpose of a complex valued matrix.

Lastly, we are also able to derive the minimal state-space representation. This system is characterized by the smallest possible dimension n_{x_2} of the state vector that is able to capture all dynamics and has the familiar state-space solution. Denote $\hat{x}_{2,t}$ the minimal state vector and $z_{2,t} := [(\hat{x}_{2,t}^f)', (\hat{x}_{2,t}^s)', (\hat{x}_{2,t}^f \otimes \hat{x}_{2,t}^f)']'$, then

$$z_{2,t} = \tilde{A}z_{2,t-1} + \tilde{B}\tilde{\xi}_t, \quad (17)$$

$$y_t = \bar{y} + \tilde{C}z_{2,t-1} + \tilde{D}\tilde{\xi}_t. \quad (18)$$

is the minimal representation of the pruned state-space. As DSGE models are based upon micro-foundations $\hat{x}_{2,t}$ is for small and medium-sized DSGE models not hard to determine.¹⁰

In summary, the pruned state-space representation is a stable system and has well-defined statistical properties, which we can exploit for our identification analysis. In particular, we see that an approximation to higher orders yields additional restrictions on the mean and variance, which may tighten identification. In section 5 it will be shown how to incorporate these additional restrictions into formal identifiability criteria and tests, but first we discuss derivatives of these objects.

4 Derivatives of solution matrices, moments and spectrum

To establish rank criteria we will need derivatives of all solution matrices, moments and spectrum with respect to the deep parameters θ . Following ideas from Iskrev (2008) and Schmitt-Grohé and Uribe (2012, Supplemental Material, Sec. A.3) we view f as well as the Jacobian of f as a function of θ and of the steady-state vector $\overline{xy}(\theta) := (\overline{x}(\theta)', \overline{y}(\theta)')'$, which is also a function of θ . Thus, implicitly we have $f(\overline{xy}(\theta), \theta) = 0$. Differentiating yields

$$df := \frac{\partial f(\overline{xy}(\theta), \theta)}{\partial \theta'} = \frac{\partial f}{\partial \overline{xy}'} \frac{\partial \overline{xy}}{\partial \theta'} + \frac{\partial f}{\partial \theta'} = 0 \quad \Leftrightarrow \quad \frac{\partial \overline{xy}}{\partial \theta'} = - \left[\frac{\partial f}{\partial \overline{xy}'} \right]^{-1} \frac{\partial f}{\partial \theta'}.$$

¹⁰For the derivation of this model representation and some practical issues regarding the minimal state vector see Appendix D.

This expression can easily be obtained analytically using e.g. MATLAB's symbolic toolbox. The derivative of the Jacobian $\mathcal{D}f(\overline{xy}(\theta), \theta)$ with respect to θ is then given by

$$d\mathcal{D}f := \frac{\partial \text{vec}(\mathcal{D}f(\overline{xy}(\theta), \theta))}{\partial \theta'} = \frac{\partial \text{vec}(\mathcal{D}f)}{\partial \overline{xy}'} \frac{\partial \overline{xy}}{\partial \theta'} + \frac{\partial \text{vec}(\mathcal{D}f)}{\partial \theta'}.$$

Note that $d\mathcal{D}f$ can be partitioned into

$$d\mathcal{D}f = \begin{pmatrix} \frac{\partial \text{vec}(\partial f(\overline{xy})/\partial x'_{t+1})}{\partial \theta'} \\ \frac{\partial \text{vec}(\partial f(\overline{xy})/\partial y'_{t+1})}{\partial \theta'} \\ \frac{\partial \text{vec}(\partial f(\overline{xy})/\partial x'_t)}{\partial \theta'} \\ \frac{\partial \text{vec}(\partial f(\overline{xy})/\partial y'_t)}{\partial \theta'} \end{pmatrix} =: \begin{pmatrix} df_1 \\ df_2 \\ df_3 \\ df_4 \end{pmatrix}.$$

This approach can be extended to calculate the analytical derivative of the Magnus-Neudecker-Hessian with respect to θ , since $H := \mathcal{H}f(\overline{xy}(\theta), \theta)$:

$$dH := \frac{\partial \text{vec}(\mathcal{H}f(\overline{xy}(\theta), \theta))}{\partial \theta'} = \frac{\partial \text{vec}(\mathcal{H}f)}{\partial \overline{xy}'} \frac{\partial \overline{xy}}{\partial \theta'} + \frac{\partial \text{vec}(\mathcal{H}f)}{\partial \theta'}.$$

Our MATLAB code writes all analytical derivatives using symbolic expressions into script files for further evaluation. For numerical derivatives we use the two-sided central difference method described in appendix C. Note that we use the following notation: $dX := \frac{\partial \text{vec}(X)}{\partial \theta'}$ for a matrix X .

Derivatives of first-order solution matrices Let $K_{n,q}$ be the commutation¹¹ matrix of order (n, q) and

$$F = -(h'_x g'_x \otimes I_{n_x+n_y})df_2 - (h'_x \otimes I_{n_x+n_y})df_1 - (g'_x \otimes I_{n_x+n_y})df_4 - df_3,$$

then the derivatives of the first-order solution matrices are given by:

$$\begin{aligned} \begin{bmatrix} dg_x \\ dh_x \end{bmatrix} &= \begin{bmatrix} (h'_x \otimes f_2) + (I_{n_x} \otimes f_4) & (I_{n_x} \otimes f_2 g_x) + (I_{n_x} \otimes f_1) \end{bmatrix}^{-1} \cdot F, \\ dg'_x &= K_{n_y, n_x} dg_x, \\ dh'_x &= K_{n_x, n_x} dh_x, \\ dh_x^t &= (I_{n_x} \otimes (h_x)^{t-1})dh_x + (h'_x \otimes I_{n_x})dh_x^{t-1}, \quad t \geq 2. \end{aligned}$$

See Schmitt-Grohé and Uribe (2012, Supplemental Material, Sec. A.3) for the derivation of these results.

Derivatives of second-order solution matrices Differentiating (3) with respect to θ requires the analytical derivatives of Q^{-1} , R , S^{-1} and T . See appendix B for the derivation of these objects. Then the analytical derivatives of the second-order solution matrices with respect to θ can be summarized

¹¹See Magnus and Neudecker (1999, p. 46) for the definition and Magnus and Neudecker (1999, p. 182) for an application regarding derivatives.

as

$$\begin{aligned} \mathbf{d} \begin{bmatrix} \text{vec}(g_{xx}) \\ \text{vec}(h_{xx}) \end{bmatrix} &= -Q^{-1} \mathbf{d}R - (\text{vec}(R)' \otimes I_{n_x^2(n_x+n_y)}) \mathbf{d}Q^{-1}, \\ \mathbf{d} \begin{bmatrix} h_{ss} \\ g_{ss} \end{bmatrix} &= -(T' \otimes I_n) \mathbf{d}S^{-1} - S^{-1} \mathbf{d}T. \end{aligned}$$

The Jacobians of H_{xx} and G_{xx} are then simple permutations of $\mathbf{d}h_{xx}$ and $\mathbf{d}g_{xx}$.

Derivatives of pruned state-space solution matrices Differentiating A , B , C , c and d with respect to θ is a straightforward application of Algorithm 1 for partitioned matrices described in appendix B. It requires the analytical derivatives of first- and second-order solution matrices (see above) as well as of $\sigma\eta$, which is given analytically by the model. Having these Jacobians it is straightforward to compute the derivatives of the minimal state-space solution matrices \tilde{A} , \tilde{B} , \tilde{C} , \tilde{c} and \tilde{d} by simply removing the entries corresponding to unnecessary states.

Derivatives of moments Differentiating the expressions for the means of z_t (10) and y_t (13) with respect to θ requires the analytical derivatives of the pruned state-space solution matrices, whereas differentiating the variance of z_t (11), the variance of y_t (14) and autocovariances of y_t (15) is straightforward using the vec -operator. The only tedious part is the derivation of $\mathbf{d}\Sigma_\xi$. Due to computational considerations we compute instead $\mathbf{d}(B\Sigma_\xi B')$ using Algorithm 1 for partitioned matrices, see appendix B for the derivation of this object. The analytical derivatives of the first two moments are then given by

$$\begin{aligned} \mathbf{d}\mu_z &= \left([(I_{2n_x+n_x^2} - A)']^{-1} \otimes (I_{2n_x+n_x^2} - A)^{-1} \right) \mathbf{d}A + (I_{2n_x+n_x^2} - A)^{-1} \mathbf{d}c, \\ \mathbf{d}\mu_y &= \mathbf{d}\bar{y} + (I \otimes C) \mathbf{d}\mu_z + (\mu'_z \otimes I) \mathbf{d}C + \mathbf{d}d, \\ \mathbf{d}\Sigma_z &= [I_{(2n_x+n_x^2)^2} - (A \otimes A)]^{-1} [(A\Sigma_z \otimes I_{2n_x+n_x^2}) \mathbf{d}A + (I_{2n_x+n_x^2} \otimes A\Sigma_z) \mathbf{d}(A') + \mathbf{d}(B\Sigma_\xi B')], \\ \mathbf{d}\Sigma_y &= (C\Sigma_z \otimes I_{n_y}) \mathbf{d}C + (C \otimes C) \mathbf{d}\Sigma_z + (I_{n_y} \otimes C\Sigma_z) \mathbf{d}(C'), \\ \mathbf{d}(\Sigma_y(t)) &= (I_{n_y} \otimes CA^t \Sigma_z) \mathbf{d}(C') + (C \otimes CA^t) \mathbf{d}\Sigma_z + (C\Sigma_z \otimes C) \mathbf{d}(A^t) + (C\Sigma_x (A')^t \otimes I_{n_y}) \mathbf{d}C. \end{aligned}$$

where we used Theorem 1 of appendix B and $\mathbf{d}(X^{-1}) = -(X')^{-1} \otimes X^{-1} \mathbf{d}X$, see Magnus and Neudecker (1999, p. 184).

Derivative of spectral density We will now show how to obtain the derivative of $\Omega_y(\omega; \theta_0)$ w.r.t. θ in equation (16) analytically. To this end, we divide the interval $[-\pi; \pi]$ into N subintervals to obtain $N + 1$ frequency indices, ω_s denotes the s -th frequency in the partition. The following steps can be done in parallel: For each $\omega_s (s = 1, \dots, N + 1)$ we first compute the derivative of $H_\xi(e^{-i\omega_s}; \theta_0)$ and its conjugate transpose using the expression in appendix B. Then we have for each ω_s

$$\mathbf{d}\Omega_y(\omega_s, \theta_0) = \frac{1}{2\pi} \left[(H_\xi^* B\Sigma_\xi B' \otimes I_{n_y}) \mathbf{d}H_\xi + (H_\xi^* \otimes H_\xi) \mathbf{d}(B\Sigma_\xi B') + (I_{n_y} \otimes H_\xi B\Sigma_\xi B') \mathbf{d}(H_\xi^*) \right].$$

5 Identification criteria based on rank conditions

Suppose that data d_t is generated by the model with parameter vector θ_0 . The criteria we will derive stem basically all from Theorem 4 in Rothenberg (1971), which essentially states identifiability conditions based on injectivity of functions. Formally, given an objective function $f(\theta)$ a sufficient condition for θ_0 being globally identified is given by

$$f(\theta_1) = f(\theta_0) \Rightarrow \theta_1 = \theta_0$$

for any $\theta_1 \in \Theta$. If this is only true for values θ in an open neighborhood of θ_0 , the identification of θ_0 is local. Since most estimation methods in the econometric DSGE literature – e.g. full-information likelihood methods or limited-information methods like impulse-response matching or GMM – exploit information from the first two moments or spectrum of data, we will focus on the mean, autocovariances and spectrum of observable variables. Since population moments are functions of data, the fundamental idea is to check, whether the mapping from θ to these population moments is unique. Then basic mathematical results for systems of equations can be applied. This set of criteria is the most basic and the closest to the ideas of the early work on identification in systems of linear equations, since it is based upon the uniqueness of a solution (Koopmans and Reiersøl 1950; Fisher 1966; Hannan 1976). Consequently, rank and order conditions are going to be derived, and it is also possible to pinpoint the (sets of) parameters that are indistinguishable.

In the literature three formal methods based on ranks have been proposed to check identification in DSGE models via (i) observational equivalent first and second moments (Iskrev 2010), (ii) observational equivalent spectral densities (Qu and Tkachenko 2012) and (iii) implications from control theory for observational equivalent minimal systems (Komunjer and Ng 2011).¹² The pruned state-space is a linear system in the extended state vector z_t and has well-defined statistical properties, a vector-moving-average as well as a minimal representation. Thus, we are able to extend all aforementioned methods based on ranks for our non-linear DSGE model by simply using the pruned state-space representation (PSS).¹³ All proofs follow the original theorems and propositions with only slight changes and modifications in model representation and assumptions; they can be requested from the author.

Proposition 1 (Iskrev PSS) *For $t = 0, 1, \dots, T - 1$ stack all theoretical first and second moments given by equations (13), (14) and (15) into a vector*

$$\bar{m}(\theta, T) := \left(\mu'_y \quad \text{vech}(\Sigma_y)' \quad \text{vec}(\Sigma_y(1))' \quad \dots \quad \text{vec}(\Sigma_y(T-1))' \right)'$$

Assume that $\bar{m}(\theta, q)$ is a continuously differentiable functions of $\theta \in \Theta$. Let $\theta_0 \in \Theta$ be a regular point, then θ is locally identifiable at a point θ_0 from the mean and autocovariogram of y_t if and only

¹²For a review and methodological comparison of these techniques in the log-linearized case see Mutschler (2014).

¹³Note that we work with a second-order approximation for illustration purposes. Higher-order approximations are straightforward extension of the proposed concepts and ideas.

if

$$\bar{M}(q) := \frac{\partial \bar{m}(\theta_0, q)}{\partial \theta'} \quad (19)$$

has a full column rank equal to the number of parameters for $q \leq T$.

Proof Follows Iskrev (2010, Theorem 2) and Rothenberg (1971, Theorem 6).

Remark In other words, we exploit Iskrev (2010)'s approach and insert the expressions for the first and second moments of the pruned state-space into a vector. The test is concerned whether these moments are uniquely determined by the deep parameters. This gives immediate rise to a necessary condition: the number of identifiable parameters does not exceed the dimension of $\bar{m}(\theta, T)$. Iskrev (2010) also proposes a necessary condition, that is checking injectivity of the mapping from the deep parameters to the solution matrices. For this, stack all elements of the mean and the solution matrices that depend on θ into a vector τ :

$$\tau(\theta) := \left(\mu'_y \quad c' \quad d' \quad \text{vec}(A)' \quad \text{vec}(C)' \quad \text{vech}(B\Sigma_\xi B')' \right)'$$

and consider the factorization $\bar{M}(q) = \frac{\partial \bar{m}(\theta, q)}{\partial \tau(\theta)'} \frac{\partial \tau(\theta)}{\partial \theta'}$. An immediate corollary implies that a point θ_0 is locally identifiable only if the rank of

$$J := \frac{\partial \tau(\theta_0)}{\partial \theta'} \quad (20)$$

at θ_0 is equal to n_θ . This condition is, however, only necessary, because τ may be unidentifiable.

Proposition 2 (Qu & Tkachenko PSS) *Assume that the spectral density in equation (16) is continuous in ω and continuous and differentiable in $\theta \in \Theta$. Let $\theta_0 \in \Theta$ be a regular point, then θ is locally identifiable at a point θ_0 from the mean and spectrum of y_t if and only if*

$$\bar{G}(\theta_0) = \int_{-\pi}^{\pi} \left(\frac{\partial \text{vec}(\Omega_y(\omega; \theta_0)')}{\partial \theta'} \right)' \left(\frac{\partial \text{vec}(\Omega_y(\omega; \theta_0))}{\partial \theta'} \right) d\omega + \frac{\partial \mu_y(\theta_0)'}{\partial \theta} \frac{\partial \mu_y(\theta_0)}{\partial \theta'}$$

is nonsingular, i.e. its rank is equal to the number of parameters.

Proof Follows Qu and Tkachenko (2012, Theorem 2).

Remark Similar to Iskrev (2010)'s approach, Qu and Tkachenko (2012) focus on the dynamic structure of the DSGE model; however, they work in the frequency domain.¹⁴ We exploit their ideas and check whether the mean and spectrum of observables is uniquely determined by the deep parameters at all frequencies using the pruned state-space representation. Note, that even when using analytical derivatives we still have to divide the interval $[-\pi; \pi]$ into sufficient subintervals N to numerically

¹⁴If the spectral density matrix is continuous there is a one-to-one relationship to the autocovariogram $\Sigma_d(j) = \int_{-\pi}^{\pi} e^{ik\omega} \Omega_d(\omega, \theta) d\omega, j = 0, \pm 1, \dots$

approximate the integral.¹⁵ That is, we can compute $\bar{G}(\theta_0)$ using

$$\bar{G}(\theta_0) \approx \frac{2\pi}{N+1} \sum_{s=1}^{N+1} \mathbf{d}\Omega_y(\omega_s, \theta_0)' \mathbf{d}\Omega_y(\omega_s, \theta_0) + \mathbf{d}\mu_y(\theta_0)' \mathbf{d}\mu_y(\theta_0).$$

The dimension of $\bar{G}(\theta_0)$ is always $n_\theta \times n_\theta$, since the spectral density matrix (16) has n_d^2 entries. Focusing on $\bar{G}(\theta_0)$ is similar to Rothenberg (1971), who looks at the Hessian of the parametric density function in the Gaussian case. In fact, for the Normal distribution both approaches are equivalent. Moreover, in the applications, we exclude the mean restrictions ($\mathbf{d}\mu_y(\theta_0)$) to check whether the parameters are identifiable only through the spectrum. We denote the corresponding matrix with $G(\theta_0)$.

Proposition 3 (Komunjer & Ng PSS) *Consider the minimal DSGE model given in equations (17) and (18). Assume that $\Lambda : \theta \mapsto \Lambda(\theta)$ is continuously differentiable on Θ with $\Lambda(\theta) := (\text{vec}(\tilde{\mathbf{A}})', \text{vec}(\tilde{\mathbf{B}})', \text{vec}(\tilde{\mathbf{C}})', \text{vec}(\tilde{\mathbf{D}})', \text{vech}(\Sigma_\xi)')$. Further denote Ξ_{n_ξ} as the left-inverse of the $n_\xi^2 + n_\xi(n_\xi + 1)/2$ duplication matrix \mathcal{G}_{n_ξ} for $\text{vech}(\Sigma_\xi)$.¹⁶ Let $\theta_0 \in \Theta$ be a regular point, then θ is locally identifiable at a point θ_0 from the mean, autocovariances and spectrum of y_t if and only if*

$$\bar{\Delta}(\theta_0) := \begin{pmatrix} \mathbf{d}\mu_d(\theta_0) & 0_{n_y \times n_{z_2}^2} & 0_{n_y \times n_\xi^2} \\ \Delta_\Lambda(\theta_0) & \Delta_T(\theta_0) & \Delta_U(\theta_0) \end{pmatrix} := \begin{pmatrix} \mathbf{d}\mu_y & 0_{n_y \times n_{z_2}^2} & 0_{n_y \times n_\xi^2} \\ \mathbf{d}\tilde{\mathbf{A}} & \tilde{\mathbf{A}}' \otimes I_{n_{z_2}} - I_{n_{z_2}} \otimes \tilde{\mathbf{A}} & 0_{n_{z_2}^2 \times n_\xi^2} \\ \mathbf{d}\tilde{\mathbf{B}} & \tilde{\mathbf{B}}' \otimes I_{n_{z_2}} & I_{n_\xi} \otimes \tilde{\mathbf{B}} \\ \mathbf{d}\tilde{\mathbf{C}} & -I_{n_{z_2}} \otimes \tilde{\mathbf{C}} & 0_{n_y n_{z_2} \times n_\xi^2} \\ \mathbf{d}\tilde{\mathbf{D}} & 0_{n_y n_\xi \times n_{z_2}^2} & I_{n_\xi} \otimes \tilde{\mathbf{D}} \\ \frac{\partial \text{vech}(\Sigma_\xi)}{\partial \theta'} & 0_{(n_\xi(n_\xi+1)/2) \times n_{z_2}^2} & -2\Xi_{n_\xi} [\Sigma_\xi \otimes I_{n_\xi}] \end{pmatrix}$$

has full column rank equal to $n_\theta + n_{z_2}^2 + n_\xi^2$.

Proof Follows Komunjer and Ng (2011, Proposition 3) and Hannan (1971, Theorem 1).

Remark Based upon results from control theory for minimal systems Komunjer and Ng (2011) derive restrictions implied by equivalent spectral densities (or equivalent autocovariances) without actually computing them as in Propositions 1 and 2.¹⁷ Intuitively, equivalent spectral densities arise if either (i) for a given size of shocks, each transfer function is potentially obtained from a multitude of quadruples of solution matrices, or (ii) there are many pairs of transfer functions and size of shocks that jointly generate the same spectral density. In $\bar{\Delta}(\theta_0)$ there are four blocks to consider: (1) The rank of the first n_y rows must equal n_θ , if the mean is uniquely determined by the deep parameters. (2) The rank of $\Delta_\Lambda(\theta_0)$ must equal n_θ , if the solution matrices are sensitive

¹⁵Regarding numerical derivatives we use the two-sided central difference method described in appendix C to compute for each ω_s the non-vectorized derivative $\frac{\partial \Omega_y(\omega_s; \theta_0)}{\partial \theta_j}$ and stack these into a big matrix. The typical element of \bar{G} is then given by $\bar{G}_{jk}(\theta) = \int_{-\pi}^{\pi} \text{tr} \left\{ \frac{\partial \Omega_y(\omega; \theta)}{\partial \theta_j} \frac{\partial \Omega_y(\omega; \theta)}{\partial \theta_k} \right\} d\omega$ which can be approximated by $\bar{G}_{jk}(\theta_0) \approx \frac{2\pi}{N+1} \sum_{s=1}^{N+1} \text{tr} \left\{ \frac{\partial \Omega_y(\omega_s; \theta_0)}{\partial \theta_j} \frac{\partial \Omega_y(\omega_s; \theta_0)}{\partial \theta_k} \right\}$ with $j, k = 1, \dots, n_\theta$.

¹⁶See Magnus and Neudecker (1999, p. 49) for the definition of the duplication matrix.

¹⁷Komunjer and Ng (2011) actually establish two conditions for identification depending on the relation between the number of shocks and observables. Here we focus on singular and squared systems ($n_\varepsilon \leq n_y$) and assume fundamental innovations. Moreover, in the commonly used squared case ($n_\varepsilon = n_y$) both conditions coincide.

to changes in parameters. (3) The rank of $\Delta_T(\theta_0)$ must equal $n_{z_2}^2$ so that the identity matrix is the only local similarity transformation. In other words, full rank of Δ_T means there exist only one quadruple generating the z-Transform for the spectral density. (4) The rank of $\Delta_U(\theta_0)$ must equal n_ξ^2 so that the spectral factorization is locally uniquely determined. Put differently, full rank of Δ_U indicates that there exist a unique pair of z-Transform and dynamic structure of the stochastic innovations that generate the spectral density. Further, we also get a necessary order condition: $n_\theta + n_{z_2}^2 + n_\xi^2 \leq n_y + (n_{z_2} + n_y)(n_{z_2} + n_\xi) + n_\xi(n_\xi + 1)/2$.

6 Implementation

All presented methods exploit the dynamic structure of the pruned solution of a non-linear DSGE model to define mappings and establish conditions for local injectivity of the mappings. For all procedures we are able to derive necessary as well as sufficient conditions for identification based on ranks of Jacobians. For calculating the ranks we use the Singular-Value-Decomposition and count the nonzero entries on the main diagonal. Obviously, this requires a specification of the tolerance level, for which we use on the one hand a range from 1e-3 to 1e-19, and on the other hand a robust tolerance level that depends on the size of the matrix. If a parameter is identified for very large tolerance levels, then it is most likely strongly identified. If it is identified only for very low levels, then it is most likely weakly identified. In the case of rank deficiency we are able to pinpoint sets of problematic parameters by analyzing the nullspace. This will be a vector of zeros, if a parameter does not affect the objective at hand. Further the columns that are linearly dependent indicate that these sets of parameters are indistinguishable. While this approach is computationally very fast, we find that in some cases there were redundancies in the subsets, since larger subsets sometimes include smaller ones. Thus, a more robust method is to consider the powerset and check the criteria for all possible subsets of parameters. In our experience this Brute-Force approach yields more reliable results and is computationally just slightly slower, because, if we find a subset of parameters that are not identified, we can exclude that subset from higher-order subsets.

There are also some further numerical issues at hand. In particular choosing the lag order T as well as the number of subintervals N for the frequencies may change results, since strictly speaking the criteria are only valid for $T, N \rightarrow \infty$. In practice, however, this is not a question of heavily sensitive results¹⁸, but rather one of speed: the higher T or N , the more time the calculations need. Komunjer and Ng (2011)'s approach is hence the fastest, since we only have to evaluate the solution matrices and their derivatives (which we also have to do for the other criteria). In this line of thought, note that all methods depend heavily on the solution matrices and suffer from possible numerical instability of the solution algorithm. However, since we use the same framework and algorithms across methods, we are able to neglect this effect in our applications in section 7.

The different interpretations of Iskrev's and Komunjer and Ng's criteria can also be used as diagnostics for model building. For instance both J as well as Δ_Λ check the mapping from the structural parameters to the pruned state-space parameters (note that J also includes the mean).

¹⁸In most practical cases T in between 10 and 100 will be sufficient, since the higher the lag the less informative the identification restrictions. Further we experienced with different values for N and find that the results hardly change. Thus an N in the order of 10000 is sufficient as well.

The evaluation yields parameters that do not influence the reduced-form and may be thus obsolete. A researcher is hence able to reparametrize the model prior to estimation. Moreover, given a known shock a rank deficient $\Delta_{\Lambda T}$ indicates that two structures (e.g. two different policies) might cause the same impulse response of the model, so we have to be careful interpreting the importance of shocks. In contrast given a rank deficient $\Delta_{\Lambda U}$ we cannot be sure, whether it is the size of the shock or a similar propagating mechanisms, that yields the same dynamic structure of the model. Qu and Tkachenko (2012)’s test does not give such diagnostics, however, their approach can be used directly for a quasi-maximum likelihood estimation in the frequency domain. Moreover, it is possible to get insight into the size of the local neighborhood of the unidentified parameters via so-called non-identification curves.

Lastly, all procedures check only local identification. Thus, one has to make sure that this procedure is valid for a sufficient range of parameters. Therefore, in our applications, we check all criteria given first a specific point (e.g. calibrated parameters or prior mean) and second given many draws from a prespecified prior domain of θ . In this way, we have a quasi-global flavor of our rank criteria for the pruned state-space.

7 Applications

7.1 The Kim (2003) model

This model extends the neoclassical growth model to include investment adjustment costs twofold: First intertemporal adjustment costs, which involve a non-linear substitution between capital and investment, are introduced into the capital accumulation equation govern by a parameter ϕ . Second multisectoral costs, which are captured by a non-linear transformation between consumption and investment, enter the budget constraint given a parameter θ . See appendix E for the model equations. In the original paper Kim (2003) log-linearizes the model and shows analytically that there is observational equivalence between these two specifications: “[W]hen a model already has a free parameter for intertemporal adjustment costs, adding another parameter for multisectoral adjustment costs does not enrich the model dynamics” (Kim 2003, p. 534). So given a first-order approximation (θ, ϕ) are observational equivalent, since they enter as a ratio $\frac{\phi+\theta}{1+\theta}$ into the solution. However, considering an approximation to the second-order yields additional restrictions on the first two moments and spectrum, as can be seen in Table 1.

[Table 1 about here.]

All criteria yield unanimously the result that (θ, ϕ) are distinguishable using a second-order approximation. This result is robust across tolerance levels as well as the choice of derivatives. A comparison of the indicators for the solution matrices, J and Δ_{Λ} , shows that the identification structure changes mainly through additional restrictions on the mean.

The same result holds when we repeat the analysis for a 100 random draws from the prior domain. For illustration purposes, we add similar to Ratto and Iskrev (2011) a parameter *dummy* into the analysis, which does not enter the model. As is evident in Figure 1(a), all criteria indicate that *dummy* and (θ, ϕ) are not identifiable in a first-order approximation. Given a second-order approximation

and using the pruned state-space criteria the situation is different: now, in all cases it is only dumpy that is not identifiable. We thus conclude that an approximation to the second order yields additional restrictions on the mean to identify θ and ϕ separately. All tests indicate that θ and ϕ are no longer observational equivalent and the model can be identified using the non-linear DSGE model. This result is – as far as we know – new to the literature.

[Figure 1 about here.]

7.2 The An and Schorfheide (2007) model

This model is a prototypical DSGE model often cited in the literature concerning lack of identification. The authors already show that (in the version we use in appendix E) the set of parameters (ν, ϕ) and the steady-state ratio $1/g = c/y$ do not enter the log-linearized solution. However, using a second-order approximation and the particle filter they conclude that “the log-likelihood is slightly sloped in $1/g = c/y$ dimension. Moreover, (...) the quadratic likelihood (...) suggests that ν and ϕ are potentially separately identifiable” (An and Schorfheide 2007, p. 164). Further, Komunjer and Ng (2011), Mutschler (2014), Ratto and Iskrev (2011) and Qu and Tkachenko (2012) show that the coefficients entering the Taylor-rule $(\psi_1, \psi_2, \rho_R, \sigma_R)$ are not separately identifiable in the log-linearized model. However, An and Schorfheide argue that “the non-linear approach is able to extract more information on the structural parameters from the data. For instance, it appears that the monetary policy parameter such as ψ_1 can be more precisely estimated with the quadratic approximation” (An and Schorfheide 2007, p. 164). We will confirm these alluring results formally by checking our rank criteria for the second-order pruned state-space. First we look at the prior mean.

[Table 2 about here.]

Table 2 shows that across criteria we are formally able to proof that indeed using a second-order approximation yields additional restrictions to identify all parameters of the model. Again a comparison between J and Δ_Λ sheds light into the structure of identification in the second-order approximation; namely, breaking with certainty-equivalence there is information through the mean that spills over to identify previously non-identifiable parameters. The same result holds when we repeat the analysis for a 100 random draws from the prior domain again including a parameter dumpy that does not enter the model at all. As can be seen in Figure 2 for a first-order approximation the Taylor-rule coefficients, (ν, ϕ) and c/y enter the problematic sets, whereas in the second-order approximation, in all cases, we are able to identify all parameters.

[Figure 2 about here.]

In summary, we confirm An and Schorfheide (2007)’s approach to estimate the model using a second-order approximation and non-linear estimation methods. Breaking with certainty-equivalence yields additional information that can be used to identify all parameters of the model. The formal proof of this feature of the non-linear model is – as far as we know – new to the literature.

8 Conclusion

We establish formal rank criteria for local identification of the deep parameters of a DSGE model in the pruned state-space system. Our procedures can be implemented prior to actually using non-linear estimation methods. In this way we show identification of the Kim (2003) and An and Schorfheide (2007) model, when solved by a second-order approximation. We want to emphasize that this result is from a theoretical standpoint: Given an infinite sample size it is theoretically possible to estimate sets of parameters which are not identifiable in the log-linearized model. The methods, however, do not point towards a specific estimation method. How good we are actually able to estimate the Kim (2003) model, given a reasonable size, is left for further research. Accordingly, An and Schorfheide (2007) show that using a particle filter only weakly identifies the model.

Even though our exposition is based on the second-order, an extension to higher-orders is straightforward, since the pruned state-space always results in a system which is linear in an accordingly extended state vector. A further extension is to establish rank criteria for other non-linear DSGE model specifications as long as we are able to calculate moments or the spectrum of the data-generating-process. For instance, Bianchi (2013) derives analytical moments for Markov-switching models, which we can use in a similar fashion to establish rank criteria for Markov-switching DSGE models.

9 References

- An, Sungbae and Frank Schorfheide (2007). “Bayesian Analysis of DSGE Models”. In: *Econometric Reviews* 26.2-4, pp. 113–172 (cit. on pp. 1, 14, 15, 22, 24).
- Anderson, Gary (2008). “Solving Linear Rational Expectations Models: A Horse Race”. In: *Computational Economics* 31.2, pp. 95–113 (cit. on p. 3).
- Andreasen, Martin M. (2011). “Non-linear DSGE models and the optimized central difference particle filter”. In: *Journal of Economic Dynamics & Control* 35.10, pp. 1671–1695 (cit. on p. 1).
- Andreasen, Martin M. (2013). “Non-linear DSGE models and the central difference Kalman filter”. In: *Journal of Applied Econometrics* 28.6, pp. 929–955 (cit. on p. 1).
- Andreasen, Martin M., Jesús Fernández-Villaverde, and Juan F. Rubio-Ramírez (2014). *The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications*. URL: <https://sites.google.com/site/mandreasendk/home-1> (visited on 05/19/2014) (cit. on pp. 1, 2, 4, 5).
- Bianchi, Francesco (2013). *Methods for Measuring Expectations and Uncertainty in Markov-Switching Models*. URL: http://http://public.econ.duke.edu/~fb36/Papers_Francesco_Bianchi.htm (visited on 09/21/2014) (cit. on p. 15).
- Binning, Andrew (2013). *Third-Order Approximation of Dynamic Models without the Use of Tensors*. URL: <http://www.norges-bank.no/en/Published/Papers/Working-Papers/2013/WP-201313/> (visited on 09/21/2014) (cit. on p. 2).
- Blanchard, Olivier J. and Charles M. Kahn (1980). “The Solution of Linear Difference Models under Rational Expectations”. In: *Econometrica* 48.5, pp. 1305–1311 (cit. on p. 3).

- Fernández-Villaverde, Jesús and Juan F. Rubio-Ramírez (2007). “Estimating Macroeconomic Models: A Likelihood Approach”. In: *Review of Economic Studies* 74.4, pp. 1059–1087 (cit. on p. 1).
- Fisher, Franklin M. (1966). *The Identification Problem in Econometrics*. New York: McGraw-Hill (cit. on p. 9).
- Gomme, Paul and Paul Klein (2011). “Second-order approximation of dynamic models without the use of tensors”. In: *Journal of Economic Dynamics and Control* 35.4, pp. 604–615 (cit. on pp. 2, 3).
- Hannan, E. J. (1971). “The Identification Problem for Multiple Equation Systems with Moving Average Errors”. In: *Econometrica* 39.5, pp. 751–765 (cit. on p. 11).
- Hannan, E. J. (1976). “The Identification and Parameterization of Armax and State Space Forms”. In: *Econometrica* 44.4, pp. 713–723 (cit. on p. 9).
- Herbst, Edward and Frank Schorfheide (2013). “Sequential Monte Carlo Sampling for DSGE Models”. In: *Journal of Applied Econometrics* Forthcoming (cit. on p. 1).
- Iskrev, Nikolay I. (2008). “Evaluating the information matrix in linearized DSGE models”. In: *Economics Letters* 99.3, pp. 607–610 (cit. on p. 6).
- Iskrev, Nikolay I. (2010). “Local identification in DSGE models”. In: *Journal of Monetary Economics* 57.2, pp. 189–202 (cit. on pp. 1, 9, 10).
- Ivashchenko, Sergey (2014). “DSGE Model Estimation on the Basis of Second-Order Approximation”. In: *Computational Economics* 43, pp. 71–82 (cit. on p. 1).
- Kim, Jinill (2003). “Functional equivalence between intertemporal and multisectoral investment adjustment costs”. In: *Journal of Economic Dynamics and Control* 27.4, pp. 533–549 (cit. on pp. 1, 13, 15, 22).
- Kim, Jinill et al. (2008). “Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models”. In: *Journal of Economic Dynamics and Control* 32.11, pp. 3397–3414 (cit. on pp. 1, 4, 21).
- Klein, Paul (2000). “Using the generalized Schur form to solve a multivariate linear rational expectations model”. In: *Journal of Economic Dynamics and Control* 24.10, pp. 1405–1423 (cit. on p. 3).
- Kollman, Robert (2014). “Tractable Latent State Filtering for Non-linear DSGE Models Using a Second-Order Approximation and Pruning”. In: *Computational Economics*, Forthcoming (cit. on p. 1).
- Komunjer, Ivana and Serena Ng (2011). “Dynamic Identification of Dynamic Stochastic General Equilibrium Models”. In: *Econometrica* 79.6, pp. 1995–2032 (cit. on pp. 1, 9, 11, 12, 14).
- Koop, Gary, M. Hashem Pesaran, and Ron P. Smith (2013). “On Identification of Bayesian DSGE Models”. In: *Journal of Business and Economic Statistics* 31.3, pp. 300–314 (cit. on p. 1).
- Koopmans, Tjalling C. and Olav Reiersøl (1950). “The Identification of Structural Characteristics”. In: *Annals of Mathematical Statistics* 21.2, pp. 165–181 (cit. on p. 9).
- Lan, Hong and Alexander Meyer-Gohde (2013). *Pruning in Perturbation DSGE Models - Guidance from Nonlinear Moving Average Approximations*. URL: <http://ideas.repec.org/p/hum/wpaper/sfb649dp2013-024.html> (cit. on p. 4).

- Lombardo, Giovanni and Harald Uhlig (2014). *A theory of pruning*. URL: <http://www.ecb.europa.eu/pub/research/working-papers/html/index.en.html> (cit. on p. 4).
- Magnus, Jan R. (2010). “On the concept of matrix derivative”. In: *Journal of Multivariate Analysis* 101.9, pp. 2200–2206 (cit. on p. 2).
- Magnus, Jan R. and Heinz Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. 2nd ed. Wiley (cit. on pp. 2, 3, 7, 8, 11, 18–20).
- Mutschler, Willi (2014). *Identification of DSGE Models - A Comparison of Methods*. URL: <http://www.mutschler.eu/identification> (visited on 05/24/2014) (cit. on pp. 9, 14).
- Pollock, D. Stephen G. (2013). “On Kronecker products, tensor products and matrix differential calculus”. In: *International Journal of Computer Mathematics*, pp. 1–15 (cit. on p. 2).
- Qu, Zhongjun and Denis Tkachenko (2012). “Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models”. In: *Quantitative Economics* 3.1, pp. 95–132 (cit. on pp. 1, 9, 10, 13, 14).
- Ratto, Marco and Nikolay I. Iskrev (2011). *Identification analysis of DSGE models with DYNARE*. URL: https://www.ifk-cfs.de/fileadmin/downloads/events/conferences/monfispol2011/RATTO_IdentifFinal.pdf (visited on 09/12/2012) (cit. on pp. 13, 14).
- Rothenberg, Thomas J. (1971). “Identification in Parametric Models”. In: *Econometrica* 39.3, pp. 577–591 (cit. on pp. 9–11).
- Schmitt-Grohé, Stephanie and Martin Uribe (2004). “Solving dynamic general equilibrium models using a second-order approximation to the policy function”. In: *Journal of Economic Dynamics and Control* 28.4, pp. 755–775 (cit. on pp. 2, 3).
- Schmitt-Grohé, Stephanie and Martin Uribe (2012). “What’s News In Business Cycles”. In: *Econometrica* 80.6 (cit. on pp. 6, 7).

A Magnus-Neudecker definition of Hessian

Define the steady state as $\bar{xy} := (\bar{x}', \bar{y}', \bar{x}', \bar{y}')' := (\bar{x}', \bar{y}')'$, then the Jacobian $Df(\bar{z})$ and Hessian $\mathcal{H}f(\bar{z})$ of f evaluated at the steady-state are defined as:

$$f(\bar{xy}) = \begin{pmatrix} f^1(\bar{xy}) \\ \vdots \\ f^n(\bar{xy}) \end{pmatrix}$$

$$Df(\bar{xy}) := \begin{pmatrix} \frac{\partial f(\bar{xy})}{\partial x'_{t+1}} & \frac{\partial f(\bar{xy})}{\partial y'_{t+1}} & \frac{\partial f(\bar{xy})}{\partial x'_t} & \frac{\partial f(\bar{xy})}{\partial y'_t} \end{pmatrix} = \begin{pmatrix} Df^1(\bar{xy}) \\ \vdots \\ Df^n(\bar{xy}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f^1(\bar{xy})}{\partial x'_{t+1}} & \frac{\partial f^1(\bar{xy})}{\partial y'_{t+1}} & \frac{\partial f^1(\bar{xy})}{\partial x'_t} & \frac{\partial f^1(\bar{xy})}{\partial y'_t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f^n(\bar{xy})}{\partial x'_{t+1}} & \frac{\partial f^n(\bar{xy})}{\partial y'_{t+1}} & \frac{\partial f^n(\bar{xy})}{\partial x'_t} & \frac{\partial f^n(\bar{xy})}{\partial y'_t} \end{pmatrix}$$

$$\mathcal{H}f(\bar{xy}) := Dvec((Df(\bar{xy}))') = \begin{pmatrix} \mathcal{H}f^1(\bar{xy}) \\ \vdots \\ \mathcal{H}f^n(\bar{xy}) \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f^1(\bar{xy})}{\partial x_{t+1} \partial x_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_{t+1} \partial y_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_{t+1} \partial x_t'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_{t+1} \partial y_t'} \\ \frac{\partial^2 f^1(\bar{xy})}{\partial y_{t+1} \partial x_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_{t+1} \partial y_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_{t+1} \partial x_t'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_{t+1} \partial y_t'} \\ \frac{\partial^2 f^1(\bar{xy})}{\partial x_t \partial x_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_t \partial y_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_t \partial x_t'} & \frac{\partial^2 f^1(\bar{xy})}{\partial x_t \partial y_t'} \\ \frac{\partial^2 f^1(\bar{xy})}{\partial y_t \partial x_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_t \partial y_{t+1}'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_t \partial x_t'} & \frac{\partial^2 f^1(\bar{xy})}{\partial y_t \partial y_t'} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f^n(\bar{xy})}{\partial x_{t+1} \partial x_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_{t+1} \partial y_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_{t+1} \partial x_t'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_{t+1} \partial y_t'} \\ \frac{\partial^2 f^n(\bar{xy})}{\partial y_{t+1} \partial x_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_{t+1} \partial y_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_{t+1} \partial x_t'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_{t+1} \partial y_t'} \\ \frac{\partial^2 f^n(\bar{xy})}{\partial x_t \partial x_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_t \partial y_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_t \partial x_t'} & \frac{\partial^2 f^n(\bar{xy})}{\partial x_t \partial y_t'} \\ \frac{\partial^2 f^n(\bar{xy})}{\partial y_t \partial x_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_t \partial y_{t+1}'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_t \partial x_t'} & \frac{\partial^2 f^n(\bar{xy})}{\partial y_t \partial y_t'} \end{pmatrix}.$$

f is of dimension $n \times 1$, the Jacobian $Df(\bar{z})$ of dimension $n \times (2n_x + 2n_y)$ and the Hessian $\mathcal{H}f(\bar{z})$ of dimension $n(2n_x + 2n_y) \times (2n_x + 2n_y)$.

B Deriving analytical derivatives

In order to calculate the derivatives of the solution matrices, we will use repeatedly the commutation matrix $K_{m,n}$ which transforms the $m \times n$ matrix A such that $K_{m,n}vec(A) = vec(A')$, and the following useful results from matrix differential calculus:

Theorem 1 (Derivative of products) *Let A be a $(m \times n)$ matrix, B a $(n \times o)$ matrix, C a $(o \times p)$ matrix and D a $(p \times q)$ matrix, then the derivative of $vec(ABCD)$ with respect to θ is given by*

$$d(ABCD) = (D'C'B' \otimes I_m) dA + (D'C' \otimes A) dB + (D' \otimes AB) dC + (I_q \otimes ABC) dD$$

Proof: Magnus and Neudecker (1999, p. 175). Note that $dX := \frac{\partial vec(X)}{\partial \theta'}$.

Theorem 2 (Derivative of Kronecker products) *Let X be a $(n \times q)$ matrix, Y a $(p \times r)$ matrix and $K_{r,n}$ the commutation matrix of order (r, n) , then the derivative of $vec(X \otimes Y)$ with respect to*

θ is given by

$$d(X \otimes Y) = (I_q \otimes K_{r,n} \otimes I_p) [(I_{n_q} \otimes \text{vec}(Y))dX + (\text{vec}(X) \otimes I_{pr})dY]$$

Proof: Magnus and Neudecker (1999, p. 185). Note that $dX := \frac{\partial \text{vec}(X)}{\partial \theta'}$.

Moreover, we will make use of the following algorithm:

Algorithm 1 (Derivative of partitioned matrix) Let X be a $(m \times n)$ matrix, that is partitioned such that $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$, with X_1 being $(m \times n_1)$ and X_2 being $(m \times n_2)$, $n = n_1 + n_2$.

1. Derive dX_1 and dX_2 ; dX_1 is of dimension $(mn_1 \times n_\theta)$ and dX_2 of dimension $(mn_2 \times n_\theta)$.
2. For $i = 1, \dots, n_\theta$
 - (a) Denote the i -th column of dX_1 and dX_2 as dX_1^i and dX_2^i respectively. dX_1^i is of dimension $(mn_1 \times 1)$ and dX_2^i of dimension $(mn_2 \times 1)$.
 - (b) Reshape dX_1^i into a $(m \times n_1)$ matrix $[dX_1^i]_{(m \times n_1)}$ and dX_2^i into a $(m \times n_2)$ matrix $[dX_2^i]_{(m \times n_2)}$.
 - (c) Store $\text{vec}(\begin{bmatrix} [dX_1^i]_{(m \times n_1)} & [dX_2^i]_{(m \times n_2)} \end{bmatrix})$ into the i -th column of a matrix dX .
3. dX is the derivative of X with respect to θ and is of dimension $(mn \times n_\theta)$.

Note that $dX := \frac{\partial \text{vec}(X)}{\partial \theta'}$.

Now we are able to derive the expressions for Q^{-1} , A , B^{-1} and C :

Derivative of Q^{-1} Notice that Q is partitioned into $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$,

$$\begin{aligned} Q_1 &= h'_x \otimes f_2 \otimes h'_x + I_{n_x} \otimes f_4 \otimes I_{n_x}, \\ Q_2 &= I_{n_x} \otimes (f_1 + f_2 g_x) \otimes I_{n_x}. \end{aligned}$$

Deriving $d(f_2 g_x)$ using Theorem 1 and mechanically applying Theorem 2 repeatedly, we obtain the derivatives dQ_1 and dQ_2 . Now we can use Algorithm 1 to compute dQ . However, we are interested in dQ^{-1} , thus in step 2(b) we also compute the derivative of the inverse using $-Q^{-1} \begin{bmatrix} [dQ_1^i] & [dQ_2^i] \end{bmatrix} Q^{-1}$ (Magnus and Neudecker 1999, p. 184) and store it in step 2(c) in the i -th column of $d(Q^{-1})$.

Derivative of R Regarding the derivative of R we first have to derive dM . This can be done in the same fashion, since M is partitioned into $M = (h_x, g_x h_x, I_{n_x}, g_x)'$. dh_x and dg_x are known, whereas $d(g_x h_x)$ can be derived using Theorem 1. Applying Algorithm 1 we get dM , whereas for the transpose we have the following relationship $dM' = K_{2(n_x+n_y), n_x} dM$. Now we are able to compute the derivative of R using Theorems 1 and 2.

Derivative of S^{-1} Since S is similarly partitioned as Q , i.e. $S = \begin{bmatrix} S_1 & S_2 \end{bmatrix}$, the derivative $d(S^{-1})$ can be calculated analogously to $d(Q^{-1})$.

Derivative of \mathbf{T} T is the sum of two matrices, for which we will derive the derivatives separately. Consider the first part, $f_2 \cdot \text{trm}[(I_{n_y} \otimes (\eta\eta'))g_{xx}]$. Since the derivatives of $(\eta\eta')$ and g_{xx} are known, it is straightforward to compute $\mathbf{d}((I_{n_y} \otimes (\eta\eta'))g_{xx})$ applying Theorems 1 and 2. The only slightly difficult part is the matrix trace function. However, Algorithm 1 can be used to overcome this difficulty. In fact, we only have one partition, for which we know the derivative. Now taking the trm of the reshaped matrix in step 2(b) and storing this in step 2(c), we get $\mathbf{d}(\text{trm}[(I_{n_y} \otimes (\eta\eta'))g_{xx}])$. Theorem 1 then yields the derivative of $f_2 \cdot \text{trm}[(I_{n_y} \otimes (\eta\eta'))g_{xx}]$. The same steps can be used to derive the derivative of the second part, $\text{trm}[(I_{n_x+n_y} \otimes N')HN(\eta\eta')]$. However, we first have to derive an expression for $\mathbf{d}N$ and $\mathbf{d}N'$. Since N is partitioned, we can use Algorithm 1 to compute $\mathbf{d}N$ and $\mathbf{d}N' = K_{2(n_x+n_y),n_x} \mathbf{d}N$.

Derivative of $B\Sigma_\xi B'$

Derivative of $H_\xi(e^{-i\omega_s})$ H_ξ is given by $C(I_{(2n_x+n_x^2)} - Ae^{-i\omega_s})^{-1}$. Closed form expressions for $\mathbf{d}C$ and $\mathbf{d}A$ are given in chapter 4 using Algorithm 1 for partitioned matrices. Thus, we only need the derivative of the inverted expression which is given by

$$\mathbf{d}((I - Ae^{-i\omega_s})^{-1}) = \left(-(I - Ae^{-i\omega_s})'^{-1} \otimes (I - Ae^{-i\omega_s})^{-1} \right) (-\mathbf{d}A \cdot e^{-i\omega_s})$$

where we used $\mathbf{d}(X^{-1}) = -(X^{-1})' \otimes X^{-1} \mathbf{d}X$, see Magnus and Neudecker (1999, p. 184). Thus, computing $\mathbf{d}H_\xi$ is a straightforward application of Theorem 1. The derivative of the conjugate transpose is given by $\mathbf{d}H_\xi^*(e^{-i\omega_s}) = K_{n_y, n_\xi} \text{conj}(\mathbf{d}H_\xi(e^{-i\omega_s}))$, where conj returns the complex conjugate.

C Deriving numerical derivatives

In order to derive the Jacobian of a function or matrix $F(\theta)$ at a point θ_0 with respect to θ , we use a two-sided finite difference method (also known as central differences). That is:

For each $j = 1, \dots, n_\theta$

1. Select a step size h_j .
2. Solve the DSGE model twice using $\bar{\theta} = \theta_0 + e_j h_j$ and $\underline{\theta} = \theta_0 - e_j h_j$ with e_j a unit vector with the j th element equal to 1.
3. Compute

$$\mathbf{d}F^j := \frac{\partial \text{vec}(F(\theta_0))}{\partial \theta_j} \approx \text{vec} \left(\frac{F(\theta_0 + e_j h_j) - F(\theta_0 - e_j h_j)}{2h_j} \right)$$

4. Store $\mathbf{d}F^j$ as the j -th column of $\mathbf{d}F$.

D Deriving the minimal state

Given the linear solution (1) and (2) of the first order approximation, we will first derive the canonical $ABCD$ -representation of the DSGE model, i.e.

...

with $z_t = ()'$ collecting all model variables and . The solution then becomes

...

Obviously, the driving force of the model is the vector of exogenous states, which we call the minimal state vector. Together with the evolution of the stochastic innovations it determines the evolution of the endogenous states, the control and the observable variables. The minimal representation is thus given by

...

Formal conditions for minimality require that for every $\theta \in \Theta$:

- (i) Controllability: For any initial state, it is always possible to design an input sequence that puts the system in the desired final state, i.e. the matrix $\begin{bmatrix} \tilde{B} & \tilde{A}\tilde{B} & \dots & \tilde{A}^{n_{x_2}-1}\tilde{B} \end{bmatrix}$ has full row rank,
- (ii) Observability: Given the evolution of the input it is always possible to reconstruct the initial state by observing the evolution of the output, i.e. the matrix $\begin{bmatrix} \tilde{C}' & \tilde{A}'\tilde{C}' & \dots & \tilde{A}^{n_{x_2}-1}'\tilde{C}' \end{bmatrix}'$ has full column rank.

Some practical issues: For small and medium-sized DSGE models the distinction between endogenous and exogenous states is given through theory, some variables are clearly endogenous (like output) and some are clearly exogenous (like technology). First, we check the rank conditions for minimality and observability given the full state vector. If the conditions fail, we remove state variables from the top until the conditions pass. Note that we remove the entries from all first-order and second-order solution matrices as well as from the derivatives corresponding to redundant state variables.

For big DSGE models the distinction of endogenous and exogenous states is often not as clear. A failsafe approach deriving the minimal state vector is to consider all possible subsets of combinations of the state vector and check the rank conditions for minimality and controllability in each case. For a different (computational) approach handling the minimal state in big DSGE models see Kim et al. (2008).

E Example Models

The Kim (2003) model

First we define an auxiliary parameter and variable:

$$s = \frac{\beta\delta\alpha}{1 - \beta + \delta\beta}, \quad \lambda_t = \frac{(1-s)^\theta}{(1+\theta)c_t^{1+\theta}}.$$

Then the model is given by the following five equations f :

$$\begin{aligned} \lambda_t(1+\theta) \left(\frac{i_t}{s}\right)^\theta \left(\frac{i_t}{\delta k_t}\right)^\phi &= \beta E_t \lambda_{t+1} \left[\alpha(1+\theta)a_t^{1+\theta} k_t^{\alpha(1+\theta)-1} + (1-\delta)(1+\theta) \left(\frac{E_t i_{t+1}}{\delta k_t}\right)^\phi \left(\frac{E_t i_{t+1}}{s}\right)^\theta \right], \\ \left[(1-s) \left(\frac{c_t}{1-s}\right)^{1+\theta} + s \left(\frac{i_t}{s}\right)^{1+\theta} \right]^{\frac{1}{1+\theta}} &= a_{t-1} k_{t-1}^\alpha, \\ k_t &= \left[\delta \left(\frac{i_t}{\delta}\right)^{1-\phi} + (1-\delta) (k_{t-1})^{1-\phi} \right]^{\frac{1}{1-\phi}}, \\ \log(a_t) &= \rho_a \log(a_{t-1}) + u_{a,t}, \\ 0 &= E_t u_{a,t+1}. \end{aligned}$$

There are two exogenous states k_t and a_t , and no endogenous states. The controls are c_t and i_t and are both observable. There is one shock on technology $u_{a,t}$ which is appended to the state vector. Further we set the perturbation parameter equal to the standard deviation of the shock on technology. Thus, given our definition and ordering of variables we have

$$\varepsilon_t := u_{a,t}, \quad x_t := (k_{t-1}, a_{t-1}, u_{a,t})', \quad y_t := (c_t, i_t)', \quad \sigma = \sigma_A, \quad \eta = \begin{pmatrix} 0 & 1 \end{pmatrix}'$$

The steady-state of this model is given by

$$a = 1, \quad k = \left(\frac{\delta}{s\alpha}\right)^{\frac{1}{\alpha-1}}, \quad i = \delta k, \quad c = (1-s) \left[\frac{(\alpha k^\alpha)^{1+\theta} - s \left(\frac{i}{s}\right)^{1+\theta}}{1-s} \right]^{\frac{1}{1+\theta}}, \quad u_a = 0.$$

We will consider identification of the parameter vector θ at the local point θ_0 :

$$\begin{aligned} \theta &= (\alpha, \beta, \delta, \theta, \phi, \rho_a, \sigma_a)', \\ \theta_0 &= (0.6, 0.99, 0.0125, 1, 2, 0.7, 1)'. \end{aligned}$$

The priors are given in Table ??.

The An and Schorfheide (2007) model

First we define auxiliary parameters:

$$\beta = \frac{1}{1 + \frac{r^{(A)}}{400}}, \quad \pi^* = 1 + \frac{\pi^{(A)}}{400}, \quad \phi = \frac{\tau(1-\nu)}{\nu\kappa\pi^{*2}}, \quad g^* = \frac{1}{(c/y)^*}.$$

Then the model f consists of thirteen equations:

$$\begin{aligned}
0 &= \frac{1-\nu}{\nu\phi\pi^{*2}} (e^{\tau c_t} - 1) - (e^{\pi_t} - 1) \left[\left(1 - \frac{1}{2\nu}\right) e^{\pi_t} + \frac{1}{2\nu} \right] + \beta (e^{E_t \pi_{t+1}} - 1) e^{-\tau E_t c_{t+1} + \tau c_t + E_t dy_{t+1} + E_t \pi_{t+1}}, \\
0 &= 1 - e^{-\tau E_t c_{t+1} + \tau c_t + R_t - \rho_z z_t - E_t \pi_{t+1}}, \quad 0 = e^{c_t - y_t} - e^{-g_t} + \frac{\phi\pi^{*2}g^*}{2} (e^{\pi_t} - 1)^2, \\
0 &= R_t - \rho_R R_{t-1} - (1 - \rho_R)\psi_1\pi_t - (1 - \rho_R)\psi_2(y_t - g_t) - u_{R,t}, \\
0 &= dy_t - y_t + y_{t-1}, \quad 0 = g_t - \rho_g g_{t-1} - u_{g,t}, \quad 0 = z_t - \rho_z z_{t-1} - u_{z,t}, \\
0 &= YGR_t - \gamma^{(Q)} - 100(dy_t + z_t), \quad 0 = INFL_t - \pi^{(A)} - 400\pi_t, \\
0 &= INT_t - \pi^{(A)} - r^{(A)} - 4\gamma^{(Q)} - 400R_t, \\
0 &= E_t u_{R,t+1}, \quad 0 = E_t u_{g,t+1}, \quad 0 = E_t u_{z,t+1}.
\end{aligned}$$

There are three exogenous states R_t, g_t and z_t , and one endogenous state variable y_t . The controls are c_t, dy_t and π_t , and the observables are $YGR_t, INFL_t$ and INT_t . There are three shocks: a monetary $u_{R,t}$, a fiscal $u_{g,t}$ and a technological shock $u_{z,t}$. Further, there are no measurement errors in the model. Thus, given our definition and ordering of variables we have

$$\begin{aligned}
\varepsilon_t &= (u_{R,t}, u_{g,t}, u_{z,t})', \quad x_t = (y_{t-1}, R_{t-1}, g_{t-1}, z_{t-1}, \varepsilon_t)', \quad D = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \\
y_t &= (c_t, dy_t, \pi_t, YGR_t, INFL_t, INT_t)'.
\end{aligned}$$

Further we set the perturbation parameter equal to the standard deviation of the shock on technology, then we have

$$\sigma = \sigma_z, \quad \eta_u = \begin{bmatrix} \sigma_R/\sigma_z & 0 & 0 \\ 0 & \sigma_g/\sigma_z & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \eta_v = [], \quad \eta_x = \begin{bmatrix} 0_{4 \times 3} \\ \eta_u \end{bmatrix}, \quad \eta_d = 0_{3 \times 3}.$$

The steady-state of this model is given by

$$\begin{aligned}
y &= R = g = z = \varepsilon = c = dy = \pi = 0, \\
YGR &= \gamma^{(Q)}, \quad INFL = \pi^{(A)}, \quad INT = \pi^{(A)} + r^{(A)} + 4\gamma^{(Q)}.
\end{aligned}$$

We will consider identification of the parameter vector θ at the local point θ_0 :

$$\begin{aligned}
\theta &= (\tau, \quad \kappa, \quad \psi_1, \quad \psi_2, \quad \rho_R, \quad \rho_g, \quad \rho_z, \quad r^{(A)}, \quad \pi^{(A)}, \quad \gamma^{(Q)}, \quad \sigma_R, \quad \sigma_g, \quad \sigma_z, \quad \nu, \quad (c/y)^*)' \\
\theta_0 &= (2, \quad 0.33, \quad 1.5, \quad 0.125, \quad 0.75, \quad 0.95, \quad 0.9, \quad 1, \quad 3.2, \quad 0.55, \quad 0.002, \quad 0.006, \quad 0.003, \quad 0.1, \quad 0.85)'.
\end{aligned}$$

Note: We could also add measurement errors in the measurement equations and extend the state vector for this additional stochastic shocks. The identification and results of this paper do not change.

Table 1: Identification analysis of the Kim (2003) model, 2nd-order approximation

Tol	<i>Iskrev</i>		<i>Komunjer/Ng</i>		<i>Qu/Tkachenko</i>	
	<i>J</i>	\overline{M}	Δ_{Λ}	$\overline{\Delta}$	<i>G</i>	\overline{G}
1e-03	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	3 (3,3,3)	3 (3,3,3)
1e-05	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	6 (6,6,6)	6 (6,6,6)
1e-07	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-09	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-11	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-13	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-15	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-17	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
1e-19	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
Robust	7 (7,7,7)	7 (7,7,7)	6 (6,6,6)	107 (107,107,107)	7 (7,7,7)	7 (7,7,7)
Require	7	7	7	107	7	7

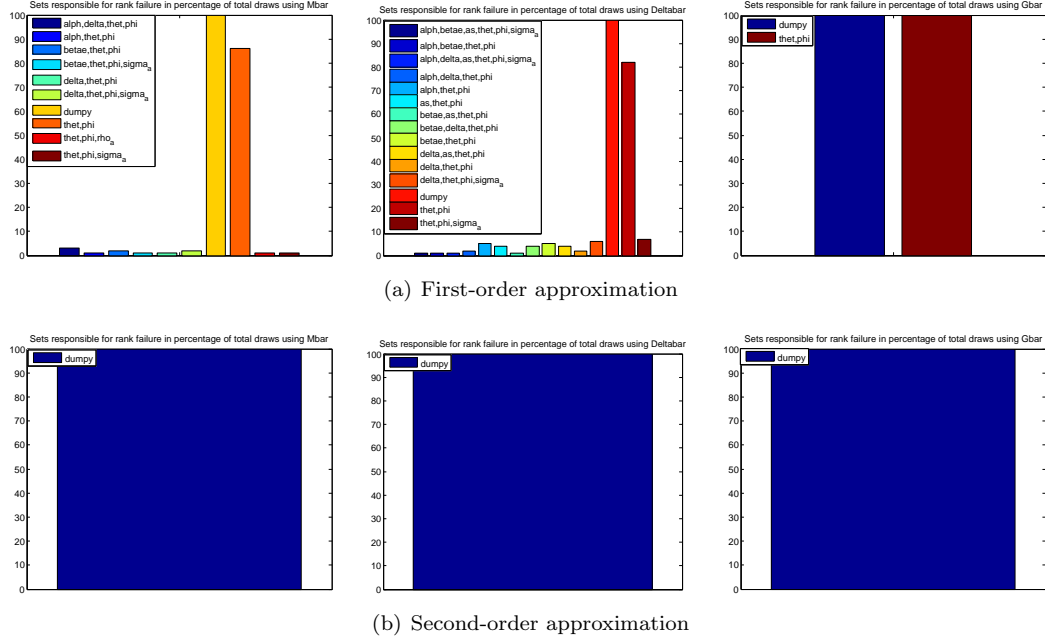
Ranks of identification tests for prior mean with analytical derivatives for different tolerance levels tol, subintervals $N = 10000$, lags in autocovariogram $T = 100$. In parenthesis are the corresponding ranks computed with numerical derivatives given differentiation steps 1e-3, 1e-7 and 1e-11, respectively.

Table 2: Identification analysis of the An and Schorfheide (2007) model, 2nd-order approximation

Tol	<i>Iskrev</i>		<i>Komunjer/Ng</i>		<i>Qu/Tkachenko</i>	
	<i>J</i>	\overline{M}	Δ_{Λ}	$\overline{\Delta}$	<i>G</i>	\overline{G}
1e-03	15 (15,15,15)	10 (9, 9, 9)	14 (14,14,15)	1140 (1140,1140,1140)	5 (5, 5, 5)	7 (7, 7, 7)
1e-05	15 (15,15,15)	15 (15,15,15)	14 (14,14,15)	1140 (1140,1140,1140)	8 (8, 8, 8)	10 (10,10,10)
1e-07	15 (15,15,15)	15 (15,15,15)	14 (14,14,15)	1140 (1140,1140,1140)	11 (11,11,11)	13 (13,13,13)
1e-09	15 (15,15,15)	15 (15,15,15)	14 (14,14,15)	1140 (1140,1140,1140)	13 (13,13,13)	14 (14,14,14)
1e-11	15 (15,15,15)	15 (15,15,15)	14 (14,14,15)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
1e-13	15 (15,15,15)	15 (15,15,15)	14 (14,15,15)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
1e-15	15 (15,15,15)	15 (15,15,15)	15 (15,15,15)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
1e-17	15 (15,15,15)	15 (15,15,15)	15 (15,15,15)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
1e-19	15 (15,15,15)	15 (15,15,15)	15 (15,15,15)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
robust	15 (15,15,15)	15 (15,15,15)	14 (14,14,14)	1140 (1140,1140,1140)	13 (13,13,13)	15 (15,15,15)
Require	15	15	15	1140	15	15

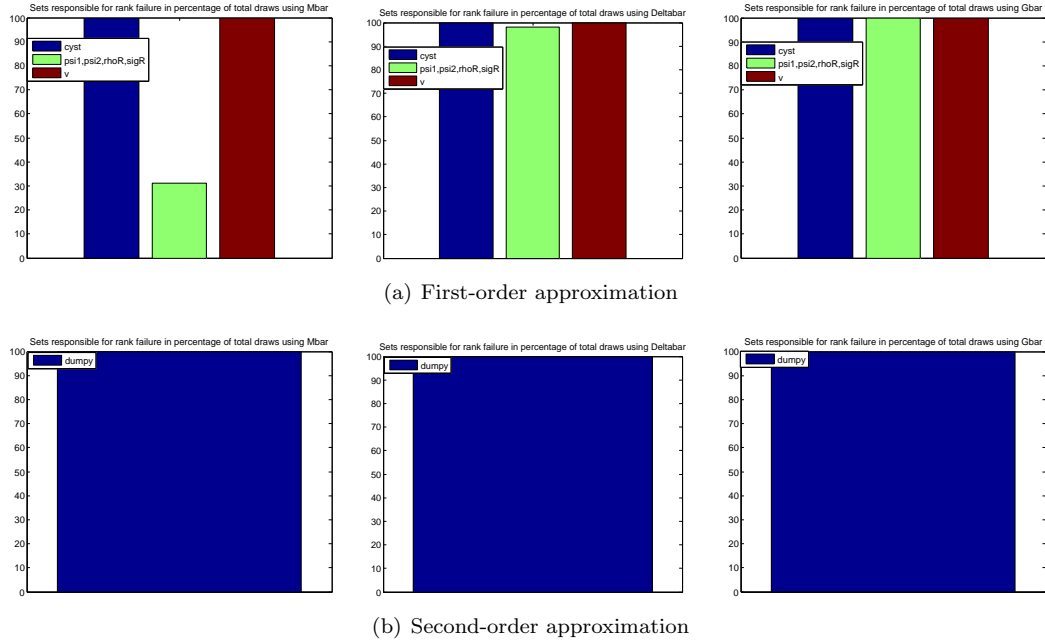
Ranks of identification tests with analytical derivatives for different tolerance levels tol, lags in autocovariogram $T = 100$, subintervals $N = 10000$. Bold indicates full rank. In parenthesis are the corresponding ranks computed with numerical derivatives given differentiation steps 1e-3, 1e-7 and 1e-11, respectively.

Figure 1: Sets responsible for non-identification in the Kim (2003) model



Identification tests for 100 draws from the prior domain using analytical derivatives with robust tolerance level, lags in autocovariogram $T = 100$, subintervals $N = 10000$.

Figure 2: Sets responsible for non-identification in the An & Schorfheide (2003) model



Identification tests for 100 draws from the prior domain using analytical derivatives with robust tolerance level, lags in autocovariogram $T = 100$, subintervals $N = 10000$.

